

Bayesian analysis of multivariate stochastic volatility with skew distribution

Jouchi Nakajima

Department of Statistical Science, Duke University

jouchi.nakajima@stat.duke.edu

December 2012

Abstract

Multivariate stochastic volatility models with skew distributions are proposed. Exploiting Cholesky stochastic volatility modeling, univariate stochastic volatility processes with leverage effect and generalized hyperbolic skew t-distributions are embedded to multivariate analysis with time-varying correlations. Bayesian prior works allow this approach to provide parsimonious skew structure and to easily scale up for high-dimensional problem. Analyses of daily stock returns are illustrated. Empirical results show that the time-varying correlations and the sparse skew structure contribute to improved prediction performance and VaR forecasts.

KEY WORDS: Generalized hyperbolic skew t-distribution; Multivariate stochastic volatility; Portfolio allocation; Skew selection; Stock returns; Value at Risk.

1 Introduction

Multivariate volatility models have attracted attention for their adaptability of variances and correlations to time series dynamics in financial econometrics in particular. A number of works discuss multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models (see e.g., Bauwens et al. 2006) and multivariate stochastic volatility (MSV) models (see e.g., Chib et al. 2006; Asai et al. 2006; Gouriéroux et al. 2009). Meanwhile, apart from symmetric distribution, several studies have addressed skew and heavy-tail properties in multivariate financial time series; their modeling strategies for return distributions use skew normal distributions (Azzalini and Valle 1996; Azzalini and Capitanio 1999, 2003; Gupta et al. 2004), a skew-

Cauchy distribution (Arnold and Beaver 2000), skew-elliptical distributions (Branco and Dey 2001; Sahu et al. 2003), and a finite mixture of skew-normal distributions (Cabral et al. 2012). (See Azzalini 2005, for a survey and discussion of skew distributions for both univariate and multivariate cases). In this context, multivariate GARCH models with skew distributions have been proposed by Bauwens and Laurent (2005) and Aas et al. (2006).

In the literature, little has been discussed about MSV models with skew error distributions. Zhang et al. (2011) develop a multivariate analysis of the generalized hyperbolic (GH) distribution with time-varying parameters driven by the score of the observation density, based on the generalized autoregressive score (GAS) model (see also, Creal et al. 2011). Ishihara et al. (2011) and Ishihara and Omori (2012) provide MSV models with a leverage effect, a stylized fact of financial returns, which induces skew conditional return distribution. In contrast, the current paper proposes MSV models with leverage effect, where structural errors follow the GH skew t -distribution. This is a natural extension of standard univariate stochastic volatility processes with skew distributions (e.g., Durham 2007; Silva et al. 2006; Nakajima and Omori 2012) to multivariate analysis; time-varying covariance components are incorporated based on the Cholesky decomposition of volatility matrices, which is increasingly used in time series analysis (e.g., Pinheiro and Bates 1996; Smith and Kohn 2002; Lopes et al. 2012). A salient feature is that prior works on a developed Bayesian approach allow for parallel computation of conditional posteriors, which enables the new model to easily scale up to higher dimensions.

Further, the new model includes a structure of skew selection for the multivariate series. Bayesian sparsity modeling has become a popular method to explore parsimonious models in a wide range of statistical analysis (see e.g., West 2003). Standard sparsity priors for variable selection in regression models (George and McCulloch 1993, 1997; Clyde and George 2004) are employed for selecting zero or non-zero skewness parameter in the GH skew t -distribution for each series. As a related work, Panagiotelis and Smith (2010) consider the sparsity prior on a coefficient of skew in a multivariate skew t -distribution. In the current paper, the sparsity prior is assumed for the skewness parameter in the GH skew t -distribution. Empirical studies using time series of stock returns show that the skewness selection, in addition to the dynamic correlated structure, reduces uncertainty of parameters and improves forecasting ability.

Section 2 defines the new MSV models with the GH skew t -distribution. Section 3 discusses Bayesian analysis and computation for model fitting. An illustrative example in Section 4 uses a time series of S&P500 Sector Indices to provide detailed evaluation of the proposed models with comparisons to standard MSV models. Section 5 presents a higher dimensional study of world-wide stock price indices to demonstrate the practical utility of the approach. Section 6 provides some summary comments.

2 Multivariate stochastic volatility and skew distribution

This section first introduces the GH skew t -distribution in a univariate case in Section 2.1, and then defines the new class of MSV models with the skew distribution in Section 2.2.

2.1 GH skew t -distribution

Suppose a univariate time series, $\{w_{it}, t = 1, 2, \dots\}$, follows the GH skew t -distribution that can be written in the form of *normal variance-mean mixture* as

$$w_{it} = m_i + \beta_i z_{it} + \sqrt{z_{it}} \varepsilon_{it}, \quad (1)$$

with $\varepsilon_{it} \sim N(0, 1)$, and $z_{it} \sim IG(\nu_i/2, \nu_i/2)$, where IG denotes the inverse gamma distribution. The previous studies often assume that $m_i = -\beta_i c_i$, where $c_i \equiv E(z_{it}) = \nu_i/(\nu_i - 2)$, for $E(w_{it}) = 0$, and $\nu_i > 4$ for the finite variance of w_{it} , which is taken here. This is a special case of a more general class of the GH distribution (see e.g., Aas and Haff 2006). As discussed by Prause (1999) and Aas and Haff (2006), the parameters of the general GH distribution are typically difficult to jointly estimate. Therefore, the current paper uses the GH skew t -distribution in the form of eqn. (1), which includes necessary parameters enough to describe the skew and heavy-tails of the financial return distributions (Nakajima and Omori 2012).

A key structure of the class of GH distributions is that the random variable is represented by the normal variance-mean mixture: a linear combination of two random variables that follow standard normal distribution, and the generalized inverse Gaussian (GIG) distribution, a more general class of the inverse gamma distribution taken for the GH skew t -distribution. The combination of a mixing weight β_i in eqn. (1), called an asymmetric parameter, and the scale parameter ν_i determines the skewness and heavy-tailedness of the resulting distribution. As illustrated by Aas and Haff (2006) and Nakajima and Omori (2012), the β_i represents the degree of skew with ν_i fixed, and the ν_i represents the degree of heavy-tails with β_i fixed. There are other definitions for the skew t -distributions in the literature (Hansen 1994; Fernández and Steel 1998; Prause 1999; Jones and Faddy 2003; Azzalini and Capitanio 2003). However, the GH skew t -distribution defined above has a great advantage in consonance with Bayesian modeling of latent variables. The representation of the normal variance-mean mixture leads to an efficient computation with conditional samplers for the latent variables in model fitting using Markov chain Monte Carlo (MCMC) methods, described in Section 3.

2.2 Cholesky multivariate stochastic volatility

Define a $k \times 1$ vector response time series $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$, ($t = 1, 2, \dots$). A standard Cholesky MSV model defines $\mathbf{y}_t \sim N(\mathbf{0}, \Sigma_t)$ with the triangular reduction $\mathbf{A}_t \Sigma_t \mathbf{A}_t' = \mathbf{\Lambda}_t^2$, where

\mathbf{A}_t is the lower triangular matrix of covariance components with unit diagonal elements and $\mathbf{\Lambda}_t$ is diagonal with positive structural variance elements: viz.

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -a_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -a_{k1,t} & \cdots & -a_{k,k-1,t} & 1 \end{pmatrix}, \quad \mathbf{\Lambda}_t = \begin{pmatrix} \lambda_{1t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_{kt} \end{pmatrix}.$$

This implies $\Sigma_t = \mathbf{A}_t^{-1} \mathbf{\Lambda}_t^2 (\mathbf{A}_t')^{-1}$, and $\mathbf{y}_t = \mathbf{A}_t^{-1} \mathbf{\Lambda}_t \mathbf{e}_t$, where $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{I})$. The construction of this Cholesky decomposition has appeared in previous works for constant covariance matrices (Pinheiro and Bates 1996; Pourahmadi 1999; Smith and Kohn 2002; George et al. 2008) and dynamic covariance modeling with stochastic volatility models (Cogley and Sargent 2005; Primiceri 2005; Lopes et al. 2012). A salient feature in Bayesian modeling of Cholesky MSV models for time-varying parameters of the covariance/variance elements is that the approach reduces the multivariate dynamics to univariate volatility processes that form a state space representation, as discussed by Lopes et al. (2012). The new idea exploits the Cholesky structure for modeling MSV and embeds the GH skew t -distribution as follows.

The new class of models is defined by

$$\mathbf{y}_t = \mathbf{A}_t^{-1} \mathbf{\Lambda}_t \mathbf{w}_t,$$

where $\mathbf{w}_t = (w_{1t}, \dots, w_{kt})'$ is the $k \times 1$ vector whose element w_{it} independently follows the GH skew t -distribution defined by eqn. (1). Define $\mathbf{h}_t = (h_{1t}, \dots, h_{kt})'$ as the $k \times 1$ vector of stochastic volatility in $\mathbf{\Lambda}_t$ with $h_{it} = \log(\lambda_{it}^2)$, for $i = 1, \dots, k$, and $\mathbf{a}_t = (a_{1t}, \dots, a_{pt})'$ as the $p \times 1$ ($p = k(k-1)/2$) vector of the strictly lower-triangular elements of \mathbf{A}_t (stacked by rows). The time-varying processes for these Cholesky parameters are specified as

$$\begin{aligned} \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \\ \mathbf{a}_{t+1} &= \boldsymbol{\mu}_a + \boldsymbol{\Phi}_a(\mathbf{a}_t - \boldsymbol{\mu}_a) + \boldsymbol{\xi}_t, \end{aligned} \tag{2}$$

and

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} \sim N \left(\mathbf{0}, \begin{bmatrix} \mathbf{I} & \mathbf{Q} & \emptyset \\ \mathbf{Q} & \mathbf{S} & \emptyset \\ \emptyset & \emptyset & \mathbf{V} \end{bmatrix} \right),$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$, and each of $(\boldsymbol{\Phi}, \boldsymbol{\Phi}_a, \boldsymbol{S}, \boldsymbol{Q}, \boldsymbol{V})$ is assumed diagonal: $\boldsymbol{\Phi} = \text{diag}(\phi_i)$, $\boldsymbol{\Phi}_a = \text{diag}(\phi_{ai})$, $\boldsymbol{S} = \text{diag}(\sigma_i^2)$, $\boldsymbol{Q} = \text{diag}(\rho_i \sigma_i)$, and $\boldsymbol{V} = \text{diag}(v_{ai}^2)$, with $|\phi_i| < 1$ and $|\phi_{aj}| < 1$, for each $i = 1, \dots, k$, and $j = 1, \dots, p$. Thus all univariate time-varying parameters follow stationary AR(1) processes. The identity $\tilde{\mathbf{y}}_t \equiv \mathbf{A}_t \mathbf{y}_t = \boldsymbol{\Lambda}_t \mathbf{w}_t$ leads to a set of univariate stochastic volatilities with the GH skew t -distribution (Nakajima and Omori 2012):

$$\tilde{y}_{it} = \{\beta_i(z_{it} - c_i) + \sqrt{z_{it}}\varepsilon_{it}\} \exp(h_{it}/2), \quad (3)$$

$$h_{i,t+1} = \mu_i + \phi_i(h_{it} - \mu_i) + \eta_{it}, \quad (4)$$

$$\begin{pmatrix} \varepsilon_{it} \\ \eta_{it} \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Omega}_i), \quad \text{and} \quad \boldsymbol{\Omega}_i = \begin{pmatrix} 1 & \rho_i \sigma_i \\ \rho_i \sigma_i & \sigma_i^2 \end{pmatrix}, \quad (5)$$

where $(\tilde{y}_{it}, \mu_i, \phi_i, \eta_{it})$ are the i -th (diagonal) elements of $(\tilde{\mathbf{y}}_t, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\eta}_t)$, respectively, for $i = 1, \dots, k$. The ρ_i measures the correlation between ε_{it} and η_{it} , which is typically negative for stock returns as the so-called leverage effect (Yu 2005; Omori et al. 2007). The class of univariate stochastic volatility models has been well studied in the literature (e.g., Jacquier et al. 2004; Kim et al. 1998; Ghysels et al. 2002; Eraker 2004; Shephard 2005; Nakajima and Omori 2009).

In the context of MSV modeling (Chib et al. 2006; Asai et al. 2006; Gouriéroux et al. 2009), the proposed model here is a natural extension of the univariate stochastic volatility model with the GH skew t -distributions embedded in the Cholesky-type multivariate structure. Most of the multivariate skew distributions and their extension to volatility models in the previous literature often have the difficulty of scaling up in dimension of responses in terms of computation of model likelihoods and parameter estimates. In contrast, an inference of the new model reduces to that of simply k univariate stochastic volatility models; this leads an efficient and fast parallel computation under conditionally independent priors as specified below.

2.3 Skew selection

As mentioned by Primiceri (2005), it is not straightforward to theoretically explore compounded processes of covariance/variance elements in the Cholesky-type covariance matrix. (See Appendix B of Nakajima (2012) for characteristics of the resulting covariance matrix process.) To understand the skew in the Cholesky MSV models, a simulation study follows.

A sample of size $T = 1,000$ and $k = 5$ is simulated according to the proposed MSV model with fixed parameter values: $\phi_i = 0.995$, $\sigma_i = 0.05$, $\rho_i = -0.5$, $\mu_i = -9$, $\nu_i = 20$, and $a_j = 0.5$, for all i, j . These values are selected following empirical studies in previous works. The value of a_j 's set here implies correlations between responses around 0.3–0.8. For the skewness parameter, four sets of values are considered: $\boldsymbol{\beta} \equiv (\beta_1, \dots, \beta_k) =$ (i) $\mathbf{0}_{1 \times 5}$, (ii) $-\mathbf{1}_{1 \times 5}$, (iii) $(-1, -1, -1, 0, 0)$,

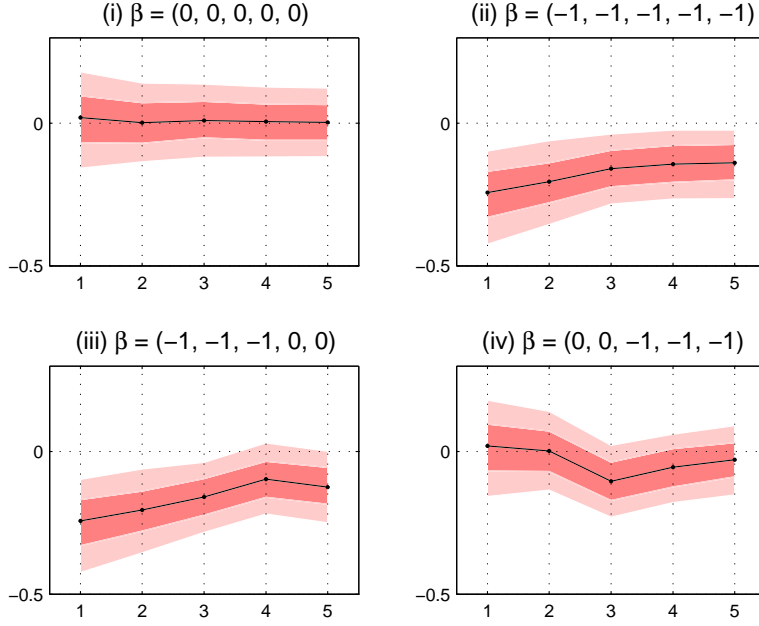


Figure 1: Skewness of simulated data: means (solid line), 50% (filled area, dark) and 80% (light) intervals for 1,000 sets of simulated series. The horizontal axis refers to the series index i .

and (iv) $(0, 0, -1, -1, -1)$. Figure 1 shows summaries of skewness of simulated data from 1,000 sets of simulation. The cases (i) and (ii) clearly exhibit no skewness and significant skewness, respectively. Interestingly, the case (iii) still yields skew observations including in the last two series ($i = 4, 5$) despite the zero skewness parameters. This is because the latter series inherit the former structural processes due to the lower triangular structure of Cholesky components (see Appendix B of Nakajima (2012)). The case (iv) confirms this mechanism; the first two series do not exhibit skewness because the corresponding skewness parameters are zero, and no inherited structure arises.

From these findings, the skewness parameter β_i 's can be redundant for the latter series in the response vector \mathbf{y}_t . Shrinkage to zero of subsets of the skewness parameters addresses skew selection in the Cholesky MSV model, exploring more parsimonious structure to reduce estimation uncertainty and improve predictions. A traditional sparsity prior for variable selection in regression models (George and McCulloch 1993, 1997; Clyde and George 2004) is employed for the skew selection. Specifically, the sparsity prior for β_i has the form

$$\beta_i \sim \kappa N(\beta_i | 0, \tau_0^2) + (1 - \kappa) \delta_0(\beta_i), \quad (6)$$

for $i = 1, \dots, k$, where δ_0 denotes the Dirac delta function at zero. This prior assigns the probability κ of taking a non-zero value and the shrinkage probability $1 - \kappa$ with a point mass at zero. Due to the structure of the normal-mean variance mixture and the conditional independence of univariate stochastic volatility processes, a conditional sampler for β_i under the sparsity prior is quite easy and simple as described in the next section.

3 Bayesian analysis and computation

Model fitting using the MCMC methods includes conditional samplers for univariate stochastic volatility models with leverage effect (Omori et al. (2007); Omori and Watanabe (2008); Nakajima and Omori (2012)) and for the state space dynamic models (e.g., Prado and West 2010). Based on observations $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ over a given time period of T intervals, the full set of latent process state parameters and model parameters in the posterior analysis are listed as follows:

- The stochastic volatility processes $h_{i,1:T}$ and mixing latent processes $z_{i,1:T}$, ($i = 1, \dots, k$);
- The covariance component process states $\mathbf{a}_{1:T}$;
- The skewness parameters β and the sparsity hyper-parameter κ ;
- Hyper-parameters defining each of the univariate stochastic volatility processes, $\theta_i \equiv \{\mu_i, \phi_i, \sigma_i, \rho_i, \nu_i\}$, ($i = 1, \dots, k$);
- Hyper-parameters defining each of the covariance component processes, $\theta_{aj} \equiv \{\mu_{aj}, \phi_{aj}, v_{aj}\}$, ($j = 1, \dots, p$).

Components of the MCMC computations are outlined as follows.

Stochastic volatility processes and mixing latent processes: The conditional posteriors for each of the latent volatility processes $h_{i,1:T}$, ($i = 1, \dots, k$) are sampled using the MCMC technique for the stochastic volatility models with leverage (Omori et al. 2007; Omori and Watanabe 2008). Nakajima and Omori (2012) implement the algorithm for the stochastic volatility with the GH skew t -distribution. Including the mixing latent process, $z_{i,1:T}$, ($i = 1, \dots, k$), these state processes are conditionally independent across i in the posteriors given all other latent variables and hyper-parameters, which allows parallel generation of the volatility processes based on eqns. (3)-(5).

Covariance component process states: Conditional on other latent process states and hyper-parameters, the MSV model reduces to a conditionally linear, Gaussian dynamic model for the

states $\mathbf{a}_{1:T}$. Specifically,

$$\hat{y}_{it} = \mathbf{a}_t^{(i)} \mathbf{x}_{it} + \hat{\varepsilon}_{it}, \quad \hat{\varepsilon}_{it} \sim N(0, z_{it}(1 - \rho_i^2)),$$

where $\hat{y}_{it} = y_{it}e^{-h_{it}/2} - \beta_i(z_{it} - c_i) - \sqrt{z_{it}\rho_i}\hat{\eta}_{it}$, $\hat{\eta}_{it} = (h_{i,t+1} - \mu_i) - \phi_i(h_{it} - \mu_i)$, $\mathbf{x}_{it} = (y_{1t}e^{-h_{1t}/2}, \dots, y_{i-1,t}e^{-h_{i-1,t}/2})'$, and $\mathbf{a}_t^{(i)}$ denotes the $1 \times (i-1)$ vector of the free parameters in the i -th row of \mathbf{A}_t , for $i = 2, \dots, k$. This observation equation is coupled with the state evolution of eqn. (2); sampling full sets of the states is implemented using the standard forward filtering, backward sampling (FFBS) algorithm (e.g., de Jong and Shephard 1995).

Skewness parameters: Conditional on all the latent states and hyper-parameters, under the prior defined by eqn. (6), the posterior for the skewness parameter β_i is given by

$$\beta_i | \cdot \sim \hat{\kappa}_i N(\beta_i | \hat{\beta}_i, \hat{\tau}_i^2) + (1 - \hat{\kappa}_i) \delta_0(\beta_i),$$

where $\hat{\beta}_i$ and $\hat{\tau}_i^2$ are the posterior mean and variance of the posterior distribution for β_i under the normal prior $N(0, \tau_0^2)$; and $\hat{\kappa}_i = \kappa b_i / (\kappa b_i + 1 - \kappa)$ with $b_i = \exp(\hat{\beta}_i^2 / 2\hat{\tau}_i^2) \hat{\tau}_i / \tau_0$. For the parameter κ , a beta prior is assumed; then we directly sample the conditional posterior given the number of β_i 's such that $\beta_i \neq 0$.

Stochastic volatility hyper-parameters: For each $i = 1, \dots, k$, traditional forms of priors for AR model parameters are assumed: normal priors for μ_i , shifted beta priors for each of (ϕ_i, ρ_i) , inverse gamma priors for σ_i^2 , and truncated gamma priors for ν_i ($\nu_i > 4$), with prior independence across i . Conditional posteriors, given the other state variables and hyper-parameters, can be sampled directly or via Metropolis-Hastings algorithms. (See Nakajima and Omori 2012, Section 2.2).

AR hyper-parameters θ_{aj} : For each $j = 1, \dots, p$, the same forms of priors are assumed for $(\mu_{aj}, \phi_{aj}, v_{aj}^2)$ as $(\mu_i, \phi_i, \sigma_i^2)$. Conditional posteriors given the states $\mathbf{a}_{1:T}$ can be sampled directly or via Metropolis-Hastings algorithms.

Note that sampling each of $(h_{i,1:T}, z_{i,1:T}, \mathbf{a}_{1:T}^{(i)}, \beta_i, \theta_i)$ can be parallelized across i . In preliminary simulation studies and the following empirical examples, MCMC streams were fairly clean and stable with quickly decaying sample autocorrelations in the same manner as the univariate stochastic volatility models.

4 A study of stock price index

The first study applies the proposed model to a series of $k = 5$ daily stock returns. An analysis particularly focuses on how the multivariate correlation mechanism and skew components re-

1	INDU	Industrials
2	CONS	Consumer Staples
3	FINL	Financials
4	ENRS	Energy
5	INFT	Information Technology

Table 1: S&P500 Sector Index. Sectors are ordered by smaller posterior means of the skewness parameter β_i obtained from univariate stochastic volatility models with the skew t -distribution.

veal dynamic relationships underlying the stock return volatilities and improve forecasting ability. Note some connections with previous work on multivariate stock return time series using dynamic volatility models (e.g. Aas et al. 2006; Chib et al. 2006; Conrad et al. 2011; Zhang et al. 2011; Ishihara and Omori 2012).

4.1 Data and model setup

The data are S&P500 Sector Indices over a time period of 1,510 business days beginning in January 2006 and ending in December 2011. The returns are computed as the log difference of the daily closing price. The series are listed in Table 1. The ordering of the series in the vector of response \mathbf{y}_t matters due to the structure based on the Cholesky decomposition. From simulation results in Section 2.3, the series are ordered by smaller posterior means of the skewness parameter β_i obtained from the univariate stochastic volatility models with the skew t -distribution. This strategy induces more parsimonious skew structure, which improves forecasting performance as discussed below.

The following priors are used: $(\phi_* + 1)/2 \sim B(20, 1.5)$ for each $* \in \{i, ai\}$, $\mu_i \sim N(-10, 1)$, $\mu_{ai} \sim N(0, 1)$, $\sigma_i^{-2} \sim G(20, 0.01)$, $v_{ai}^{-2} \sim G(20, 0.01)$, $(\rho_i + 1)/2 \sim B(1, 1)$, $\nu_i \sim G(16, 0.8)I[\nu_i > 4]$, $\beta_i \sim \kappa N(\beta_i|0, 10) + (1 - \kappa)\delta_0(\beta_i)$, and $\kappa \sim B(2, 2)$, where B and G denotes the beta and gamma distributions, respectively. The MCMC analysis was run for a burn-in period of 5,000 samples prior to saving the following 50,000 samples for posterior inferences.

The study provides forecasting performance in comparison among different specifications in the proposed class of models. The following five models are considered:

- **S:** Skew t -distribution, no sparsity on β_i ($\kappa \equiv 1$), no correlation ($\mathbf{A}_t \equiv \mathbf{I}$);
- **SS:** Skew t -distribution with sparsity on β_i , no correlation;
- **C:** Symmetric t -distribution ($\beta_i \equiv 0$), with correlation;
- **CS:** Skew t -distribution, no sparsity on β_i , with correlation;
- **CSS:** Skew t -distribution with sparsity on β_i , and correlation.

The key focus here is on the skew in return distribution, sparsity structure on the skewness parameter, and the Cholesky-type correlation mechanism in the MSV.

4.2 Forecasting performance and comparisons

Out-of-sample forecast performance is examined to compare the competing models in predicting 1 to 5 business days ahead. Forecasts are based on a posterior predictive density sampled every MCMC iteration. An experiment is implemented in a traditional recursive forecasting format; the full MCMC analysis is fit to each model to obtain the 5-horizon forecasts given data from the start of January 2006 up to business day T_i with $T_i = T_{i-1} + 5$. Specifically, each model is first estimated based on data $\mathbf{y}_{1:T_1}$ where $T_1 = 1,010$. The resulting out-of-sample predictive distributions are simulated over the following 5 business days, $t = T_1 + 1, \dots, T_1 + 5$. Next, the analysis moves ahead 5 business days to observe the next 5 observations $\mathbf{y}_{T_1+1:T_1+5}$ and reruns the MCMC based on the updated data $\mathbf{y}_{1:T_2}$, where $T_2 = T_1 + 5$, forecasting the following 5 business days $t = T_2 + 1, \dots, T_2 + 5$. This is repeated with $T_i = T_{i-1} + 5$ for $i = 2, \dots, 100$, generating a series of out-of-sample forecasts over 500 business days. This experiment allows us to explore forecasting performance over nearly 2-year periods of different financial market circumstances and so examine robustness to time periods of the prediction ability.

The first measure of formal model assessments is out-of-sample predictive densities. The log predictive density ratio (LPDR) for forecasting d business days ahead from the day t is $\text{LPDR}_t(d) = \log\{p_{M_1}(\mathbf{y}_{t+d}|\mathbf{y}_{1:t})/p_{M_0}(\mathbf{y}_{t+d}|\mathbf{y}_{1:t})\}$, where $p_M(\mathbf{y}_{t+d}|\mathbf{y}_{1:t})$ is the predictive density under model M . This quantity represents relative forecasting accuracy in the prediction exercise. Table 2 reports the LPDRs of four competing models relative to Model S at each horizon. Improvements in out-of-sample predictions are practically evident for the proposed multivariate skew models. The LPDRs for Models C and CS show relevance of correlated structure, and differences in those for Models C and CS indicate dominance of the skew component in the multivariate stock returns. The LPDRs for Model SS and comparisons in those for Models CS and CSS show that the sparsity on the skew parameters contributes to improved predictions, robustly across horizons. The LPDRs for Models CS and CSS at the 2nd and 4th horizons are

Model	Horizon (d days)					Total
	1	2	3	4	5	
SS	7.0	13.2	11.2	16.7	15.4	63.5
C	34.0	54.7	45.5	52.8	50.9	237.9
CS	88.5	216.8	85.2	240.2	144.5	775.1
CSS	92.8	230.6	91.5	259.7	156.9	831.4

Table 2: Cumulative log predictive density ratios $\text{LPDR}_t(d)$ relative to Model S.

relatively inflated, which is due to two time points under market shocks related to the European sovereign-debt crisis. In turbulent situations, the skew and correlated structures yield substantially increased improvements. Even if these two times are removed from the full period of comparison, the models still show relevant dominance over the standard MSV models.

The second measure of the formal model comparisons is based on Value-at-Risk (VaR) forecasts of portfolio returns. Using samples from the posterior predictive distribution, optimal portfolios are implemented under several allocation rules, and the VaR forecast of the resulting portfolio is obtained at each time $T_i + 1, \dots, T_i + 5$, for $i = 1 \dots, 100$. Note that Bauwens and Laurent (2005) illustrate a similar procedure in evaluation of the VaR forecasts. A main focus here is on an impact of the proposed multivariate skew model on forecasting accuracy, in particular for a tail risk of multivariate responses.

The analysis uses standard Bayesian mean-variance optimization (Markowitz 1959). Based on the samples from the posterior predictive distribution, the forecast mean vector and variance matrix of \mathbf{y}_t , denoted by \mathbf{g}_t and \mathbf{D}_t respectively, are computed. Investments are allocated according to a vector of portfolio weights, denoted by $\boldsymbol{\omega}_t$, optimized by the following allocation rule. The realized portfolio return at time t is $\boldsymbol{\omega}_t' \mathbf{y}_t$. Given a (scalar) return target m , we optimize the portfolio weights $\boldsymbol{\omega}_t$ by minimizing the forecast variance of the portfolio return among the restricted portfolios whose expectation is equal to m . Specifically, we minimize an ex-ante portfolio variance $\boldsymbol{\omega}_t' \mathbf{D}_t \boldsymbol{\omega}_t$, subject to $\boldsymbol{\omega}_t' \mathbf{g}_t = m$, and $\boldsymbol{\omega}_t' \mathbf{1} = 1$, i.e., the total sum invested on each business day is fixed. The solution is $\boldsymbol{\omega}_t^{(m)} = \mathbf{K}_t (\mathbf{1}' \mathbf{K}_t \mathbf{q}_t \mathbf{g}_t - \mathbf{g}_t' \mathbf{K}_t \mathbf{q}_t \mathbf{1})$, where $\mathbf{q}_t = (\mathbf{1}m - \mathbf{g}_t)/d_t$, and $d_t = (\mathbf{1}' \mathbf{K}_t \mathbf{1})(\mathbf{g}_t' \mathbf{K}_t \mathbf{g}_t) - (\mathbf{1}' \mathbf{K}_t \mathbf{g}_t)^2$, with $\mathbf{K}_t = \mathbf{D}_t^{-1}$. The study also considers the target-free minimum-variance portfolio given by $\boldsymbol{\omega}_t^* = \mathbf{K}_t \mathbf{1} / (\mathbf{1}' \mathbf{K}_t \mathbf{1})$. The portfolio is reallocated on each business day based on 1- to 5- business day ahead forecasts. This experiment assumes a practical situation that investors allocate their resource every business day based on weekly-updated forecasts. Note that the resources are assumed freely reallocated to arbitrary long or short positions without any transaction cost.

In summary of the VAR forecasts, the number of VaR violations, denoted by n , is counted over $N = 500$ experiment days. The expected number of violations for α quantile is αN ; under the null hypothesis that the expected ratio of violations is equal to α , the likelihood ratio statistic,

$$2 \log \left\{ \left(\frac{n}{N} \right)^n \left(1 - \frac{n}{N} \right)^{N-n} \right\} - 2 \log \{ \alpha^n (1 - \alpha)^{N-n} \},$$

is asymptotically distributed as $\chi^2(1)$ (see Kupiec 1995). Table 3 reports the number of VaR violations and results of the likelihood ratio test for $\alpha = 0.5\%$, 1% , and 5% levels, based on a range of daily target returns of $m = 0.005\%$, 0.01% , and 0.02% , implying a yearly return of

(1) Violations					
Model	α	Target return (m)			Target -free
		0.005%	0.01%	0.02%	
S	0.5%	27	18	12	37
	1%	37	26	12	50
	5%	67	53	30	79
SS	0.5%	31	20	7	41
	1%	43	32	11	48
	5%	69	60	41	79
C	0.5%	30	22	12	45
	1%	39	29	18	52
	5%	64	48	33	82
CS	0.5%	2	2	2	5
	1%	5	7	3	8
	5%	18	19	17	25
CSS	0.5%	3	2	1	4
	1%	6	5	2	6
	5%	24	26	22	24

(2) p -values					
Model	α	Target return (m)			Target -free
		0.005%	0.01%	0.02%	
S	0.5%	0.00	0.00	0.00	0.00
	1%	0.00	0.00	0.01	0.00
	5%	0.00	0.00	0.32	0.00
SS	0.5%	0.00	0.00	0.02	0.00
	1%	0.00	0.00	0.02	0.00
	5%	0.00	0.00	0.00	0.00
C	0.5%	0.00	0.00	0.00	0.00
	1%	0.00	0.00	0.00	0.00
	5%	0.00	0.00	0.12	0.00
CS	0.5%	0.74	0.74	0.74	0.16
	1%	1.00	0.40	0.33	0.22
	5%	0.13	0.20	0.08	1.00
CSS	0.5%	0.76	0.74	0.28	0.38
	1%	0.66	1.00	0.13	0.66
	5%	0.84	0.84	0.53	0.84

Table 3: VaR results: the number of violations and p -values for the likelihood ratio test. The null hypothesis is that the expected ratio of violations is equal to α .

approximately 1.25%, 2.5%, and 5.0% respectively, as well as the target-free portfolio. For a 5% significance level, the null hypothesis is rejected in almost all cases of VaR quantiles and portfolio rules for Models S, SS, and C. The large number of their VaR violations indicate that these models forecast smaller values of VaR (in their absolute value) than necessary. This optimistic risk forecast is due to lack of structure including both skewness and correlations. In contrast, the null hypothesis is not rejected in all cases for Models CS and CSS except in only one case, $\alpha = 5\%$ with $m = 0.02\%$ for Model CS.

The results from the out-of-sample forecasting experiments reveal that the skew and correlated multivariate structure contributes to the forecasting performance in terms of the predictive density and the VaR risk analysis. In particular, the sparse skew model with correlation, Model CSS, achieved the best posterior predictive densities and passed the VaR likelihood ratio test for all the realistic situations. These findings are similar to those obtained from different prior densities; a prior sensitivity analysis is provided in the Appendix.

Regarding the ordering of the response in \mathbf{y}_t , Bayesian prior works on the orderings can be considered (see Nakajima and Watanabe 2011, for reversible-jump MCMC methods in the Cholesky MSV models), although this is beyond the scope of this paper. Instead, the reverse ordering of the responses is examined here to compare with the current baseline ordering. Forecasting performances are computed for Model CSS; results show weaker forecasting ability of the reverse ordering than the baseline ordering in terms of the predictive density and VaR forecasts. This confirms that the baseline ordering based on the posterior means of β_i 's has an advantage over the reverse ordering.

4.3 Summaries of posterior inferences

Posterior estimates are summarized for results of the MSV models fit to data $\mathbf{y}_{1:T}$ with $T = T_1 = 1,010$. Figure 2 displays posteriors of model parameters for Model CSS. One remarkable finding is that the posterior for β_1 is estimated negative considerably apart from zero, although the posteriors for other β_i 's ($i = 2, \dots, 5$) exhibit shrinkage at zero; their posterior probabilities of shrinkage are about 91–94%. This parsimonious skew structure evidently improves forecasting ability compared to the non-sparsity model as reported in the previous subsection. Figure 3 plots the posterior estimates of β_i for four competing models. For Models S the β_i 's are estimated negative with reported credible intervals that are mostly apart from zero. For Model SS, moderate shrinkages are found for (β_3, β_4) , and considerable shrinkage is observed for β_{5t} . In contrast, Model CS exhibits credible intervals including zero except for β_1 , and the evident shrinkages in Model CSS yield the parsimonious skew structure.

The posteriors for the other parameters in Figure 2 are consistent with previous studies. The

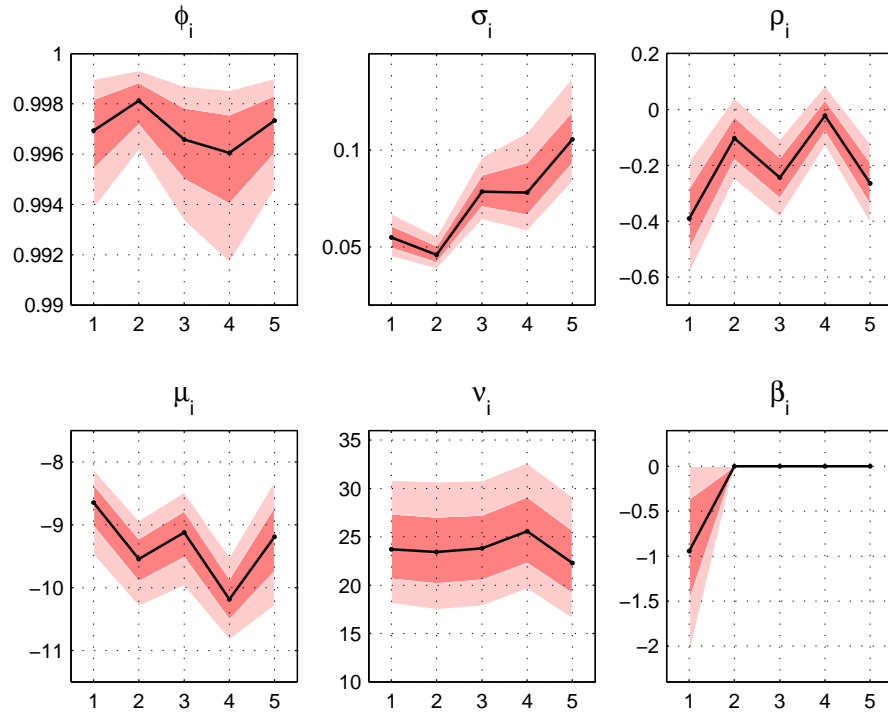


Figure 2: Posterior estimates for parameters from Model CSS: Posterior medians (solid line) and 50% (filled area, dark) and 80% (light) credible intervals. The horizontal axis refers to the series index i .

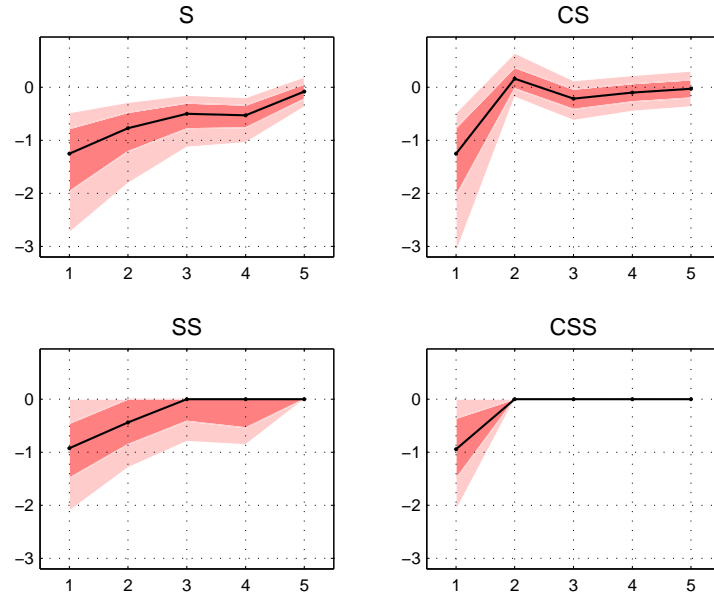


Figure 3: Posterior estimates for β_i : Posterior medians (solid line) and 50% (filled area, dark) and 80% (light) credible intervals. The horizontal axis refers to the series index i .

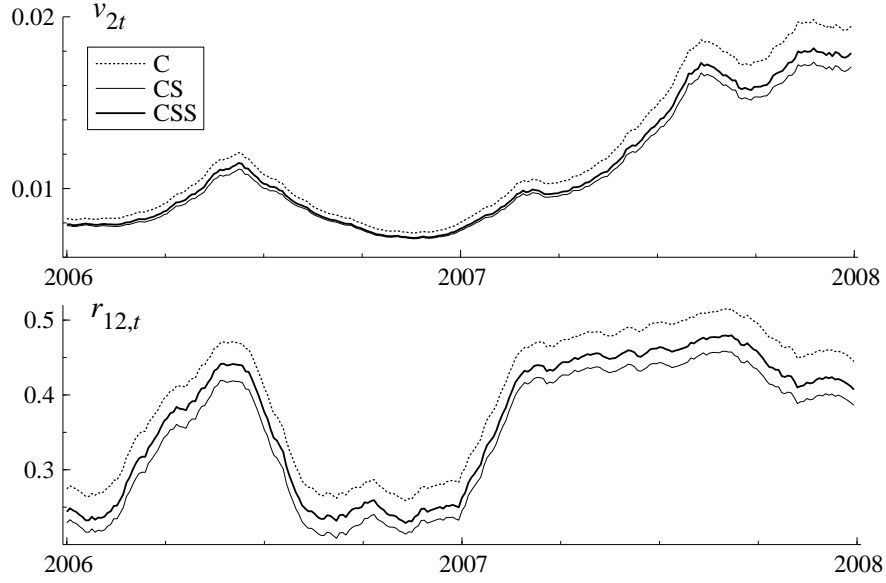


Figure 4: Posterior means of the selected standard deviation (top, CONS) and correlation (bottom, INDU-CONS) in Σ_t .

posterior medians of ρ_i 's are estimated negative, indicating the leverage effect for stock return dynamics. One possible extension from the current model is sparsity for ρ_i 's, although a result from Model CSS with the same sparsity prior embedded to ρ_i 's showed only little evidence of sparsity for all i .

Figure 4 graphs trajectories of posterior means of a selected standard deviation and correlation, denoted by v_i and r_{ij} respectively, in the resulting covariance matrix $\Sigma_t = \mathbf{A}_t^{-1} \mathbf{\Lambda}_t^2 (\mathbf{A}_t')^{-1}$. Note that the figure shows only a part of sample periods for visual clarity. The top panel shows the standard deviation of CONS ($i = 2$) from three MSV models. Model C yields higher standard deviations than the other skewed models due to the symmetric t -distribution that estimates the left tail lighter than the skew models. Model CSS yields higher standard deviations than Model CS because of the shrinkage toward zero. These differences tend to be larger in high-volatility periods. The same feature is found in the correlations; the bottom panel shows the correlation between INDU ($i = 1$) and CONS. Model CS yields less correlated structure due to its skew error distribution. Meanwhile, across the series and sample periods, the correlation is evidently time-varying for the stock return data, which results in the contribution of the Cholesky-based time-varying correlation structure to the improved prediction ability.

Further, Figure 5 shows approximated posterior joint predictive densities of $(y_{1,T+1}, y_{2,T+1})$ and $(y_{2,T+1}, y_{3,T+1})$ in surface plots with tail behaviors displayed in scatter plots. Compared to

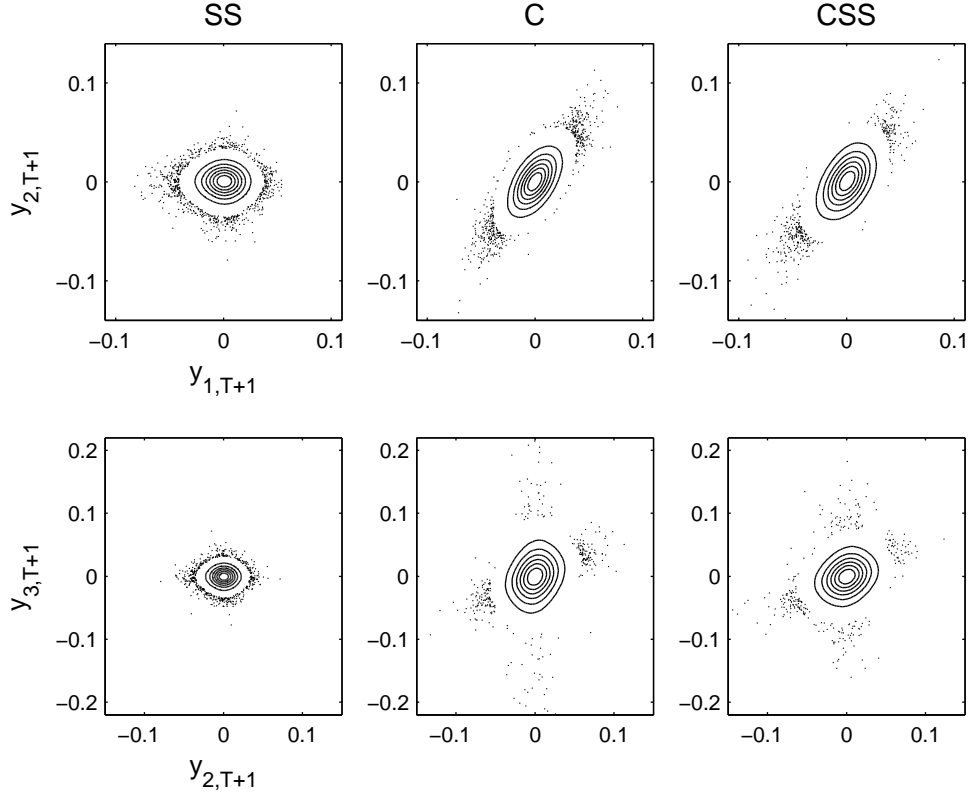


Figure 5: Surface plots for smoothed joint predictive density for $(y_{1,T+1}, y_{2,T+1})$ (top) and $(y_{2,T+1}, y_{3,T+1})$ (bottom) based on MCMC samples. Scatter plots are displayed for tail samples, defined as samples in regions where the smoothed density is less than 1% of the maximum density.

Model SS, the correlated MSV models (C and CSS) exhibit a clear image of correlated predictive densities. Model CSS yields more tail samples in the left tails due to the negative skewness. These differences result in the large improvement of VaR forecasts illustrated in the previous subsection.

5 A higher-dimensional study: World-wide stock price indices

This section provides a higher-dimensional example for the skew and correlated MSV models using $k = 20$ world-wide stock price indices (see the list of countries and regions in Table 4). These are selected as major indices traded in the global financial market; note that both the Euro and several European countries are included, although their time series do not exhibit severely high correlations. The time period is $T = 1,258$ business days beginning in January 2006 and ending in December 2010. The returns are computed as the log difference of prices at the closing time of the US market. The variables in \mathbf{y}_t are ordered by posterior means of the

1	Euro	11	Brazil
2	US	12	Spain
3	India	13	Russia
4	Taiwan	14	Swiss
5	Netherlands	15	Hong Kong
6	Japan	16	UK
7	Mexico	17	Australia
8	Sweden	18	Germany
9	France	19	Canada
10	Italy	20	Korea

Table 4: World-wide stock price indices. Countries are ordered by smaller posterior means of the skewness parameter β_i obtained from univariate stochastic volatility models with the skew t -distribution.

skewness parameter β_i obtained from the same pre-analysis, and the study uses the same prior specifications as in the previous section.

Figure 6 shows posteriors of the model parameters for Model CSS. A remarkable evidence is considerable shrinkage of β_i 's except for the first two series, suggesting much parsimonious structure induced by the skew selection. Figure 7 plots posteriors of β_i for Models S and CS. Note that the series are ordered by posterior means of β_i 's based on Model S, although the posterior medians displayed here are not monotonically increasing. Model CS exhibits interesting estimates; a posterior distribution of β_5 leans to positive, presumably for adjusting the skewness in connection with the former series on the Cholesky-type compound processes. This finding and the evidence of shrinkage in Model CSS suggest that the skewness parameter can be redundant in the correlated MSV models.

Other parameters in Figure 6 show some differences in behaviors of stock price indices among the countries. The series of Spain ($i = 12$) and Russia ($i = 13$) exhibit smaller ϕ_i 's and larger σ_i 's, implying less persistent volatility dynamics. The series of EUR ($i = 1$) and US ($i = 2$) show large leverage effects with posterior medians of ρ_i below -0.5 . An important advantage here is that the proposed Cholesky MSV models easily scales up in its dimension with the reduction of posterior computation to univariate stochastic volatility analysis.

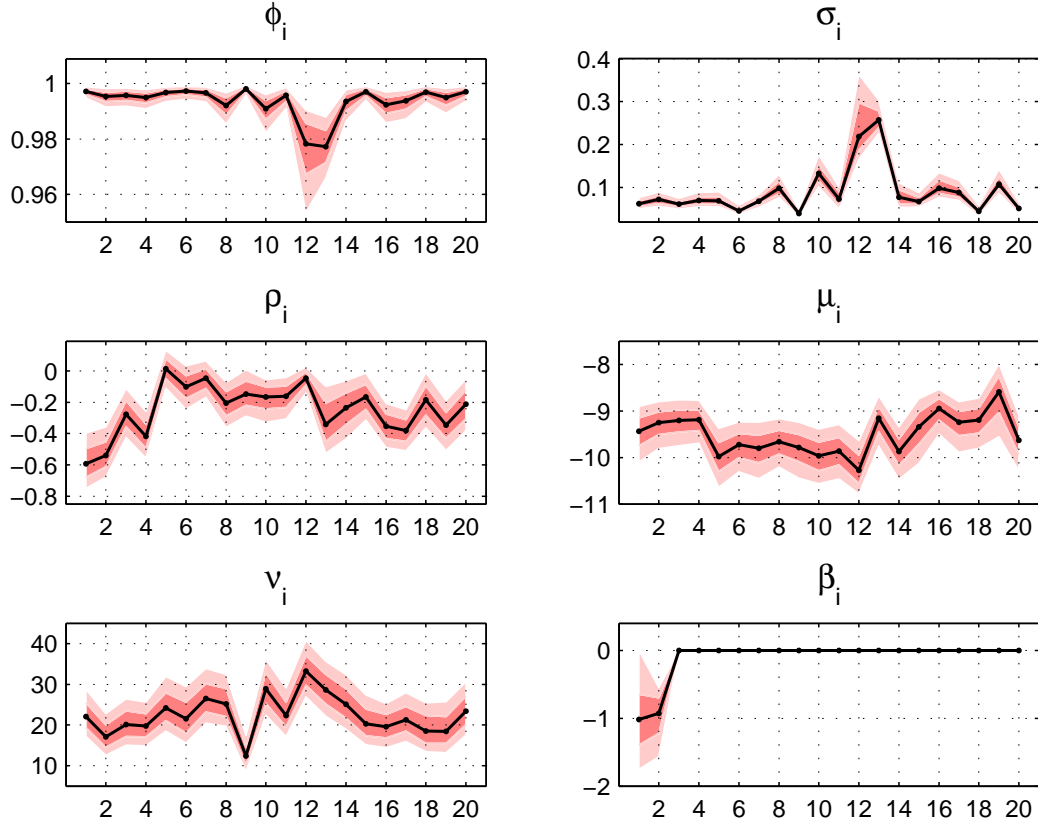


Figure 6: Posterior estimates for parameters from Model CSS: Posterior medians (solid line) and 50% (filled area, dark) and 80% (light) credible intervals. The horizontal axis refers to the series index i .

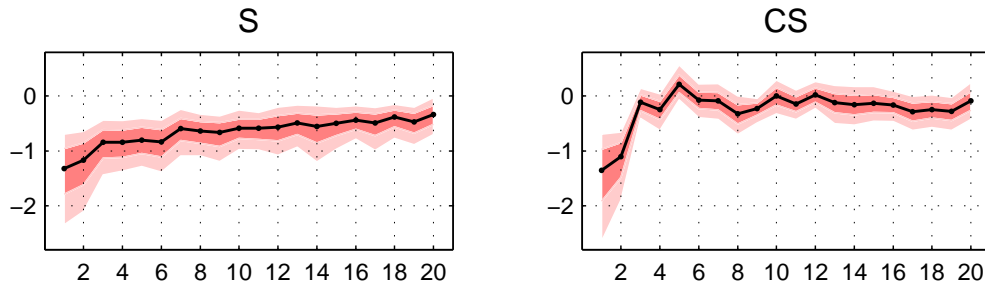


Figure 7: Posterior estimates for β_i : Posterior medians (solid line) and 50% (filled area, dark) and 80% (light) credible intervals. The horizontal axis refers to the series index i .

6 Concluding remarks

A new framework of building correlated multivariate stochastic volatility models with skew distributions is developed. The approach of Cholesky-type covariance structure effectively embeds the univariate stochastic volatility with leverage effects and the GH skew t -distributions to the multivariate analysis. The salient feature of the proposed model is the skew selection based on the sparsity prior on the skewness parameters. In stock return analyses, the empirical evidence shows the sparse skew and dynamic correlated structures contribute to improved prediction ability in terms of the predictive density and portfolio VaR forecasts, which is practically relevant to business and policy uses of such models in investment and risk management.

There are a number of methodological and computational areas for further investigation. In terms of modeling strategy, the sparse skew structure can be applied to factor stochastic volatility models, which have been widely studied in literature (Geweke and Zhou 1996; Pitt and Shephard 1999; Aguilar and West 2000; Chib et al. 2006). Also, the time-varying sparsity technique using latent threshold models proposed by Nakajima and West (2012a,b) can be employed to explore more parsimonious covariance structure for the skew MSV models. One important open question is a potential computational strategy of sequential particle learning algorithms (Carvalho et al. 2010) for the proposed MSV models, which would be useful in real-time decision making context.

Acknowledgements

The author thanks Kaoru Irie for helpful discussion. Computations were implemented using Ox (Doornik 2006).

Appendix. Prior sensitivity analysis

A forecasting study with different prior distributions is examined for the S&P500 Sector Indices data used in Section 4. Consider the following priors: (Prior-1) $\kappa \sim B(2, 8)$, (Prior-2) $\beta_i \sim \kappa N(\beta_i | -1, 2) + (1 - \kappa)\delta_0(\beta_i)$, and (Prior-3) $\nu_i \sim G(24, 0.6)I[\nu_i > 4]$. All the other parameters remain the same as the baseline priors specified in Section 4.1. Compared to the baseline priors, the new priors imply more concentrated densities of κ (Prior-1) with posterior mean 0.2, shifted from 0.5 in the baseline, β_i (Prior-2) with posterior mean -1 from 0, and ν_i (Prior-3) with posterior mean 40 from 20.

Table 5 reports the cumulative log predictive density ratios of Models CS and CSS relative to Model S, which shows little difference among the different priors. In addition, the VaR forecasts are also computed in the same way as in Section 4.2; the likelihood ratio tests indicate the null

Model	Prior	Horizon (d days)					Total
		1	2	3	4	5	
C	(1)	34.0	54.7	45.5	52.8	50.9	237.9
	(2)	34.6	54.0	44.1	50.9	50.2	233.8
	(3)	48.6	63.1	58.3	63.6	68.1	301.7
CS	(1)	88.5	217.6	84.7	240.6	143.9	775.2
	(2)	89.9	223.5	81.5	241.5	148.1	784.6
	(3)	101.6	222.7	97.0	250.9	162.6	834.7
CSS	(1)	96.8	232.1	94.3	256.8	158.6	838.6
	(2)	95.7	227.1	89.2	246.5	150.0	808.4
	(3)	117.2	245.6	116.3	276.2	181.6	937.0

Table 5: Prior sensitivity analysis: Cumulative log predictive density ratios $LPDR_t(d)$ relative to Model S.

hypothesis that the expected ratio of violations is equal to α is not rejected for Models CS and CSS with the those priors in any case considered in Table 3 at the 5% significance level. These findings indicate that the results of forecasting performance improved by the skew and correlated structure in the MSV models are quite robust regardless of prior specifications for those key hyper-parameters.

References

- Aas, K. and Haff, I. H. (2006), “The generalized hyperbolic skew Student’s t-distribution,” *Journal of Financial Econometrics*, 4, 275–309.
- Aas, K., Haff, I. H., and Dimakos, X. K. (2006), “Risk estimation using the multivariate normal inverse Gaussian distribution,” *Journal of Risk*, 8, 39–60.
- Aguilar, O. and West, M. (2000), “Bayesian dynamic factor models and portfolio allocation,” *Journal of Business and Economic Statistics*, 18, 338–357.
- Arnold, B. C. and Beaver, R. J. (2000), “The skew-Cauchy distribution,” *Statistics & Probability Letters*, 49, 285–290.
- Asai, M., McAleer, M., and Yu, J. (2006), “Multivariate stochastic volatility: A review,” *Econometric Reviews*, 25, 145–175.
- Azzalini, A. (2005), “The skew-normal distribution and related multivariate families,” *Scandinavian Journal of Statistics*, 32, 159–188.

- Azzalini, A. and Capitanio, A. (1999), “Statistical applications of the multivariate skew normal distribution,” *Journal of the Royal Statistical Society B*, 61, 579–602.
- (2003), “Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution,” *Journal of the Royal Statistical Society B*, 65, 367–389.
- Azzalini, A. and Valle, A. D. (1996), “The multivariate skew-normal distribution,” *Biometrika*, 83, 715–726.
- Bauwens, L. and Laurent, S. (2005), “A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models,” *Journal of Business & Economic Statistics*, 23, 346–354.
- Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006), “Multivariate GARCH models: A survey,” *Journal of Applied Econometrics*, 21, 70–109.
- Branco, M. D. and Dey, D. K. (2001), “A general class of multivariate skew-elliptical distributions,” *Journal of Multivariate Analysis*, 79, 99–113.
- Cabral, C. R. B., Lachos, V. H., and Prates, M. O. (2012), “Multivariate mixture modeling using skew-normal independent distributions,” *Computational Statistics & Data Analysis*, 56, 126–142.
- Carvalho, C. M., Johannes, M. S., Lopes, H. F., and Polson, N. G. (2010), “Particle learning and smoothing,” *Statistical Science*, 25, 88–106.
- Chib, S., Nardari, F., and Shephard, N. (2006), “Analysis of high dimensional multivariate stochastic volatility models,” *Journal of Econometrics*, 134, 341–371.
- Clyde, M. and George, E. I. (2004), “Model uncertainty,” *Statistical Science*, 19, 81–94.
- Cogley, T. and Sargent, T. J. (2005), “Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.” *Review of Economic Dynamics*, 8, 262–302.
- Conrad, C., Karanasos, M., and Zeng, N. (2011), “Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study,” *Journal of Empirical Finance*, 18, 147–159.
- Creal, D., Koopman, S. J., and Lucas, A. (2011), “A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations,” *Journal of Business & Economic Statistics*, 29, 552–563.

- de Jong, P. and Shephard, N. (1995), “The simulation smoother for time series models,” *Biometrika*, 82, 339–350.
- Doornik, J. (2006), *Ox: Object Oriented Matrix Programming*, London: Timberlake Consultants Press.
- Durham, G. B. (2007), “SV mixture models with application to S&P 500 index returns,” *Journal of Financial Economics*, 85, 822–856.
- Eraker, B. (2004), “Do equity prices and volatility jump? Reconciling evidence from spot and option prices,” *Journal of Finance*, 59, 1367–1403.
- Fernández, C. and Steel, M. F. J. (1998), “On Bayesian modeling of fat tails and skewness,” *Journal of the American Statistical Association*, 93, 359–371.
- George, E. I. and McCulloch, R. E. (1993), “Variable selection via Gibbs sampling,” *Journal of the American Statistical Association*, 88, 881–889.
- (1997), “Approaches for Bayesian variable selection,” *Statistica Sinica*, 7, 339–373.
- George, E. I., Sun, D., and Ni, S. (2008), “Bayesian stochastic search for VAR model restrictions,” *Journal of Econometrics*, 142, 553–580.
- Geweke, J. F. and Zhou, G. (1996), “Measuring the pricing error of the arbitrage pricing theory,” *Review of Financial Studies*, 9, 557–587.
- Ghysels, E., Harvey, A. C., and Renault, E. (2002), “Stochastic volatility,” in *Statistical Methods in Finance*, eds. Rao, C. R. and Maddala, G. S., Amsterdam: North-Holland, pp. 119–191.
- Gouriéroux, C., Jasiak, J., and Sufana, R. (2009), “The Wishart autoregressive process of multivariate stochastic volatility,” *Journal of Econometrics*, 150, 167–181.
- Gupta, A. K., González-Farías, G., and Domínguez-Molina, J. A. (2004), “A multivariate skew normal distribution,” *Journal of Multivariate Analysis*, 89, 181–190.
- Hansen, B. E. (1994), “Autoregressive conditional density estimation,” *International Economic Review*, 35, 705–730.
- Ishihara, T. and Omori, Y. (2012), “Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors,” *Computational Statistics and Data Analysis*, 56, 3674–3689.

- Ishihara, T., Omori, Y., and Asai, M. (2011), “Matrix exponential stochastic volatility with cross leverage,” Discussion paper series, CIRJE-F-812, Faculty of Economics, University of Tokyo.
- Jacquier, E., Polson, N., and Rossi, P. (2004), “Bayesian analysis of stochastic volatility with fat-tails and correlated errors,” *Journal of Econometrics*, 122, 185–212.
- Jones, M. C. and Faddy, M. J. (2003), “A skew extension of the t -distribution, with application,” *Journal of Royal Statistical Society, Series B*, 65, 159–174.
- Kim, S., Shephard, N., and Chib, S. (1998), “Stochastic volatility: Likelihood inference and comparison with ARCH models,” *Review of Economic Studies*, 65, 361–393.
- Kupiec, P. (1995), “Techniques for verifying the accuracy of risk measurement models,” *Journal of Derivatives*, 2, 173–184.
- Lopes, H. F., McCulloch, R. E., and Tsay, R. (2012), “Cholesky stochastic volatility models for high-dimensional time series,” Tech. rep., University of Chicago, Booth Business School.
- Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, New York: John Wiley and Sons.
- Nakajima, J. (2012), “Bayesian Analysis of Latent Threshold Models,” Ph.D. thesis, Duke University, Durham, N.C.
- Nakajima, J. and Omori, Y. (2009), “Leverage, heavy-tails and correlated jumps in stochastic volatility models,” *Computational Statistics and Data Analysis*, 53, 2535–2553.
- (2012), “Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student’s t -distribution,” *Computational Statistics and Data Analysis*, 56, 3690–3704.
- Nakajima, J. and Watanabe, T. (2011), “Bayesian analysis of time-varying parameter vector autoregressive model with the ordering of variables for the Japanese economy and monetary policy,” Global COE Hi-Stat Discussion Paper Series 196, Hitotsubashi University.
- Nakajima, J. and West, M. (2012a), “Bayesian analysis of latent threshold dynamic models,” *Journal of Business and Economic Statistics*, forthcoming.
- (2012b), “Dynamic factor volatility modeling: A Bayesian latent threshold approach,” *Journal of Financial Econometrics*, doi:10.1093/jfinec/nbs013.

- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), “Stochastic volatility with leverage: fast likelihood inference,” *Journal of Econometrics*, 140, 425–449.
- Omori, Y. and Watanabe, T. (2008), “Block sampler and posterior mode estimation for asymmetric stochastic volatility models,” *Computational Statistics and Data Analysis*, 52, 2892–2910.
- Panagiotelis, A. and Smith, M. (2010), “Bayesian skew selection for multivariate models,” *Computational Statistics & Data Analysis*, 54, 1824–1839.
- Pinheiro, J. C. and Bates, D. M. (1996), “Unconstrained parametrizations for variance-covariance matrices,” *Statistics and Computing*, 6, 289–296.
- Pitt, M. and Shephard, N. (1999), “Time varying covariances: A factor stochastic volatility approach (with discussion),” in *Bayesian Statistics VI*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford University Press, pp. 547–570.
- Pourahmadi, M. (1999), “Joint mean-covariance models with application to longitudinal data: Unconstrained parametrisation,” *Biometrika*, 86, 677–690.
- Prado, R. and West, M. (2010), *Time Series Modeling, Computation, and Inference*, New York: Chapman & Hall/CRC.
- Prause, K. (1999), “The Generalized Hyperbolic models: Estimation, financial derivatives and risk measurement,” PhD dissertation, University of Freiburg.
- Primiceri, G. E. (2005), “Time varying structural vector autoregressions and monetary policy,” *Review of Economic Studies*, 72, 821–852.
- Sahu, S. K., Dey, D. K., and Branco, M. D. (2003), “A new class of multivariate skew distributions with applications to Bayesian regression models,” *Canadian Journal of Statistics*, 31, 129–150.
- Shephard, N. (2005), *Stochastic Volatility: Selected Readings*, Oxford: Oxford University Press.
- Silva, R. S., Lopes, H. F., and Migon, H. S. (2006), “The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models,” *Brazilian Journal of Probability and Statistics*, 20, 67–91.
- Smith, M. and Kohn, R. (2002), “Parsimonious covariance matrix estimation for longitudinal data,” *Journal of the American Statistical Association*, 97, 1141–1153.

- West, M. (2003), “Bayesian factor regression models in the “large p, small n” paradigm,” in *Bayesian Statistics 7*, eds. Bernardo, J., Bayarri, M., Berger, J., David, A., Heckerman, D., Smith, A., and West, M., Oxford, pp. 723–732.
- Yu, J. (2005), “On leverage in a stochastic volatility model,” *Journal of Econometrics*, 127, 165–178.
- Zhang, X., Creal, D., Koopman, S. J., and Lucas, A. (2011), “Modeling dynamic volatilities and correlations under skewness and fat tails,” Duisenberg School of Finance - Tinbergen Institute Discussion Paper, No.11-078/2/DSF22.